Defense Technical Information Center Compilation Part Notice

ADP023083

TITLE: V & V of MuSES 6.0 Status Report

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Proceedings of the Ground Target Modeling and Validation Conference [13th] Held in Houghton, MI on 5-8 August 2002

To order the complete compilation report, use: ADA459530

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP023075 thru ADP023108

V & V of MuSES 6.0 Status Report

Teresa Gonda
TARDEC
Warren, MI 48397-5000

Al Curran, Alice Gerhardt, David Less, Pete Rynes, John S. Curlee
ThermoAnalytics 94X Airport Road
Calumet, MI 49913-0066

Bob Baratono
Keweenaw Research Center
1400 Townsend Drive, Houghton MI 49931

ABSTRACT

MuSES 6.0 was released late in 2001. TARDEC has undertaken a formal verification and validation procedure as described in the Army Modeling and Simulation Office Pamphlet 5-11 and as described in the Defense Modeling and Simulation Offices Guidelines on the subject. The full report is expected to be finished in the Fall of 2002. This paper will give examples of the latest results of the full environment Cubi tests and additional verification and validation tests. In addition, the paper will discuss some of the lessons learned on the V&V of engineering level models.

INTRODUCTION

The idea that everyone wants V&V is hard to dispute. What's more important however is that the DoD demands it. The Defense Modeling and Simulation Office (DMSO) released a comprehensive guide on the subject in 1996[1] and is currently updating it. In addition, they have established a Verification, Validation, and Accreditation (VV&A) technical working group to continue to mature the process, educate the community, and establish clear common nomenclature.

The Army Modeling and Simulation Office has followed suit and published guidelines under pamphlet PAM-5-11.[2] The official definitions of verification, validation, and accreditation in PAM-511 are as follows:

- "Verification is the process that determines the M&S functions as it was originally conceived, specified, and designed.
- Validation is the process that addresses the credibility of the M&S in its depiction of the modeled world.
- Accreditation is the M&S application sponsor's determination of the suitability and acceptability of the M&S to the application"

What VV&A can do is enhance a simulation's credibility and reduce the risk of its use in a particular application. What it cannot do is "guarantee that the modeling and simulation results will be correct, guarantee that the results will be correctly analyzed and interpreted, guarantee that the right model was chosen to solve the problem."[3]

UPHILL BATTLE

Even though program managers and engineers want validated models, very few organizations understand the process or want to go through the process. The reason is twofold. Firstly, it seems like an immense amount of work and secondly, there is "Miller's Law". Certainly, the impact of picking up the thick Recommended Practices Guide on VV&A is to create a sense of being overwhelmed, but truthfully, the bulk of V&V it is just documenting good software development practices that should already be going on along the way. These procedures include creating requirements specifications, showing how requirements are tracked, documenting the sanity checks (verifications) and assumptions being made along the way, tracking configuration management procedures, and then finally the validation tests that any responsible proponent would perform on any model before they used it to their satisfaction. Once all of this is documented along with the strengths and the weaknesses of the model, then others can use this V&V report to judge whether the M&S is suitable for a particular problem. This "judgement" is the accreditation. Just because an M&S has weaknesses, does not make it invalid. If the weaknesses are understood and the M&S can still answer a certain question, then the M&S can be deemed appropriate and accredited for that purpose. Figure 1 lists some of the elements set aside for V&V in MuSES.

Assumptions made	Radiation variables	Signature inputs and outputs
Overall conduction	Environmental variables	Post processor features.
Overall radiation	All three modes of heat transfer	Geometry functions
Convection variables	Analysis of inputs and displays	Graphical features
Overall Convection	Editor functions	

Figure 2 Elements targeted for verification in MuSES

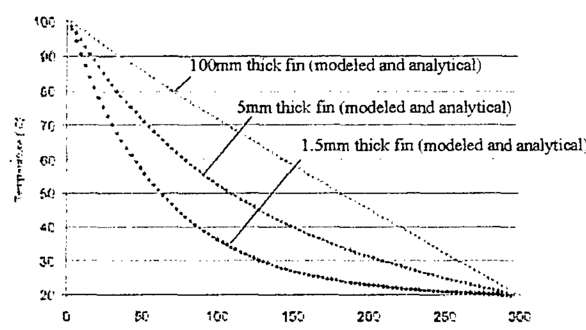

Figure 1 Results of conduction and convection validations

MILLER'S LAW

MuSES 6.0 has undergone the steps of individual V&V for the above elements. Comparisons to textbook analyses have shown agreement and the model behaves as expected. The final stage we are undertaking is face validation or output validation, which in this case is comparison to field trials using thermocouples and infrared imagers. This is the part of the process that most of those outside the V&V community think of as "validation". In the recommended procedures guide this part plays a small part in the over all process, but it is what most engineers and program managers wish to see. Unfortunately, this is also where Miller's Law comes into play.

Miller's Law states: Model *comparison to measured data is unfair*[4]

A portion of the sentiment behind this "law" is echoed in the DMSO guide. The very first principle listed in the guide is "There is no such thing as an absolutely valid model[1]". This principle addresses the idea that some might hold the mistaken belief that once a model was "validated" it was likely to give absolute answers that were "perfect". While there is a distinct wish to see plots of exact matches of modeled and measured data, under a great many circumstances that is unrealistic and not because a model is "bad". In fact, in certain cases, the modeled answer can actually be "more valid" than the measured due to uncertainties in measurements. This is the idea put forth in the adage, "no one believes a model except the one who ran it and everyone believes a measurement except the one who took it".

The reasoning behind Miller's Law is the following:

- Measurements are not always exact (due to inherent instrumentation uncertainties)
- Modelers are not always exact
- Modeling theories are often based on Measurements (see first bullet)
- Nature is not always well understood

Table 1 shows the reason as to why measurements are not always exact--in particular with respect to measuring "nature" or weather and environmental effects.

	Modeled – Measured Temperature Difference				
	Average	**Avg Abs**	**St Dev**	**Min**	**Max**
Baseline	-0.53	0.98	1.36	-10.6	6.64
Air T + 0.7 °C	0.061	0.86	1.31	-10.1	6.33
Solar Radiance + 5%	-0.26	0.91	1.29	-9.28	8.81
Solar Absorptivity + 0.05	-0.27	0.93	1.34	-9.94	8.56
Emissivity – 0.05	-0.39	0.93	1.36	-10.5	7.39
Cp 461 -> 420	-0.53	0.97	1.33	-10.4	6.72
Wind Speed – 5%	-0.49	0.99	1.35	-10.5	6.98
Air T + 1.0 °C	0.22	0.99	1.36	-9.84	7.02
Sky Radiance + 5%	0.16	1.00	1.42	-10.2	7.97
Thermal Conduct 52 ->73	-0.58	1.01	1.39	-10.8	6.32
Rotate Back 5 Deg	-0.61	1.02	1.36	-10.6	5.90
Insulated Back	-0.57	1.01	1.33	-10.3	7.82
One Layer with Air Inside	-0.62	1.05	1.50	-11.4	5.09

Table 1: Weather File Sensitivity Analysis, Output Validation

Miller's Law addresses the fact that in nature, there are many variables that fluctuate in nearly random fashion (with wind being the most dramatic) and the instruments that are used to measure nature and provide input data to models during validation exercises are not sensitive enough to capture these fluctuations. When this input data to the model is not accurate, it is unrealistic (unfair) to expect that models based on this input data will exactly predict what was then observed in the field.

Sometimes accurate data is theoretically possible, but very difficult to obtain in practice and determining <u>how</u> accurate data must be may be difficult to define. MuSES relies on material properties such as the emissivity and absorptivity broken out by wavelengths. This data set can be quite large, which can be a problem, but the bigger problem is that the data is hard to come by unless one owns the instrumentation to measure these properties. The measurement devices are expensive and the measurement itself is tedious--therefore there does not exist a large database of this type of information readily available. But the real problem with material properties is much more complex than this. The modeler needs to answer the question "what surface condition do I wish to model?" A lab measurement of a coating may be what is available for input into the model, but if one wishes to compare the results of the model to a measurement of a vehicle that has been in use for two years, the comparison could be very poor. Scratched and weather-beaten CARC paint for instance, has different properties than a coupon that is freshly painted and measured in a lab.

Miller's Law shines the light on the fact that face validation, while desirable, is a very difficult task to do well. The truth is that face validation has value and even if graphs of measured versus modeled do not match point for point, the conclusion is not automatically that there is a problem with the model. All modelers need to understand the assumptions made within the model and assumptions made about measurements used as input to the model. That said, we have proceeded forward with output validation (face validation) in the final stages of the V&V of MuSES. We will discuss the results, the examples of Miller's Law encountered, and the additional experiment designed to reduce its impact

OUTPUT VALIDATION

Since absolute comparison is not necessarily what is expected in output validation, how then can one measure a sense of "goodness?" When comparing measured and modeled data for MuSES, agreement is expected to fall within certain error bars calculated based on the above uncertainties. In addition, in these trials, we look at and compare temperature profiles over time. The results are time-stepped. Therefore, it can be said that it is also appropriate to look for agreement within plus or minus a given time step as well. Finally, there is "trend agreement." One can also visually tell if the predicted trend matches that of the measured trend. Under most circumstances, average agreement within plus or minus 2 degrees with good trend agreement is very good. Larger isolated deviations in and of themselves may not be a problem and are looked at on a case-by-case basis. Once modeling error, instrumentation error, and input data error is eliminated in these cases, algorithms and assumptions are analyzed. This is discussed in more detail later in this section.

Cubi

The geometry used in the output validation exercises of the MuSES/RadTherm code is called Cubi. Technion, in Israel first introduced this geometry in 1995 in order to validate PRISM. It was later adopted by the UK MoD. Liking the design, we chose the same geometry and dimension values in order to exchange data with our international partners. Cubi was purposefully designed such that it is fairly simple to model and troubleshoot, yet still presents some challenges to a model in a validation experiment. Firstly, past model validation experiments have shown that there is usually very good agreement in thicker armor. Therefore, in order to tax the model, thinner plates were used (which traditionally have been more difficult to match in predictions). Being box-like, terrain and sky effects are individually taxed/isolated via the different faces of the box. Then finally, the notch or step in the geometry allows faces to interact with each other (self-shadowing/self radiating).

Cubi is one meter tall by one half meter wide and one meter deep. It is constructed of 4.0-mm steel and was painted with a green CARC paint. Paint coupons were fabricated at the same time and sent out for material property measurements. Thermocouples were applied to the inside surface of the metal, and one-inch thick polystyrene foam board was pushed into place inside Cubi. The most recent thermocouple locations can be seen in the figure below.

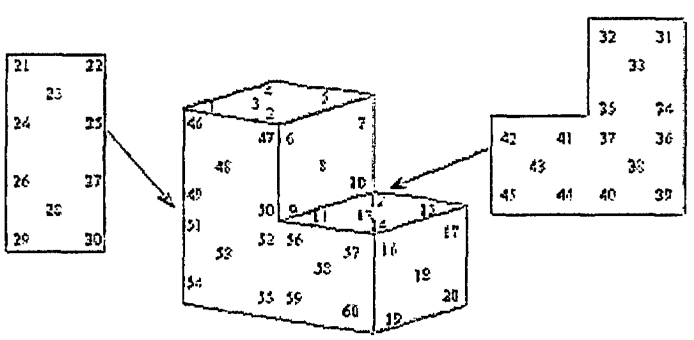

Figure 3 Cubi geometry and thermocouple map

Previous to this map, an initial test was performed in the spring in a small field close to the buildings at TACOM. There was a different thermocouple placement at that time. The results are in the figure below.

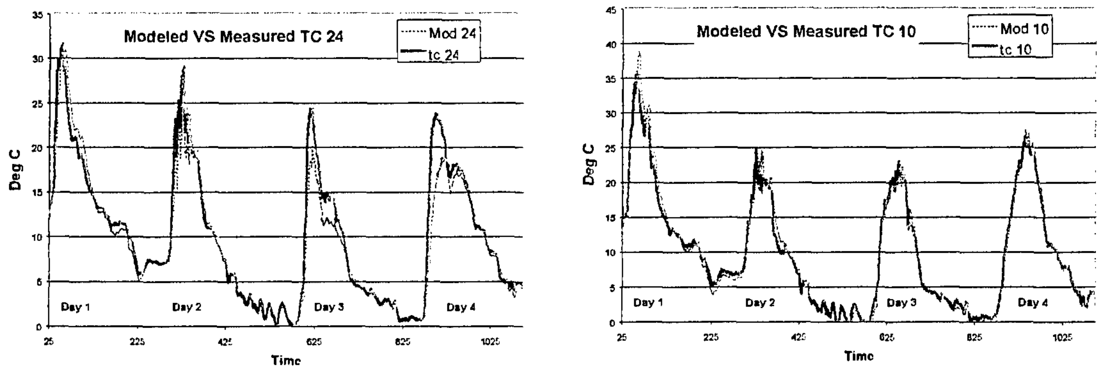

Figure 4 CUBI -- Old Cubi runs of upper east face (24) and of lower step facing up (10)

In this older trial thermocouple (tc) 24 is in the location of tc8 in figure 3 and tc10 is in the location of tc13. There are four days of consecutive comparisons. There is considerable agreement throughout the period with some problem areas during solar peak hours. A more rigorous second test was then designed to rule out testing errors and a new location was chosen that had less environmental clutter or interference from buildings and trailers that were not part of the experiment.

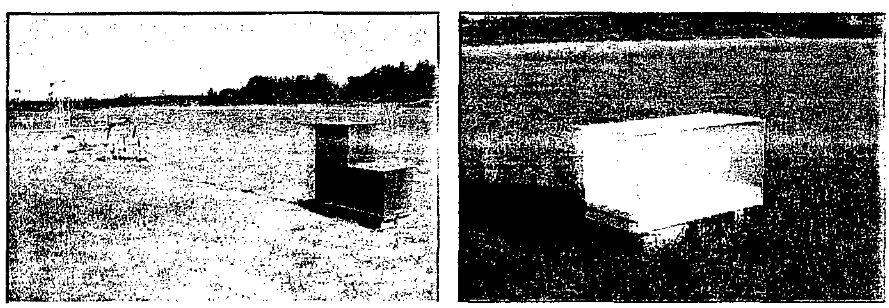

Figure 5 Cubi in the field and Plexiglas box for wind elimination experiment

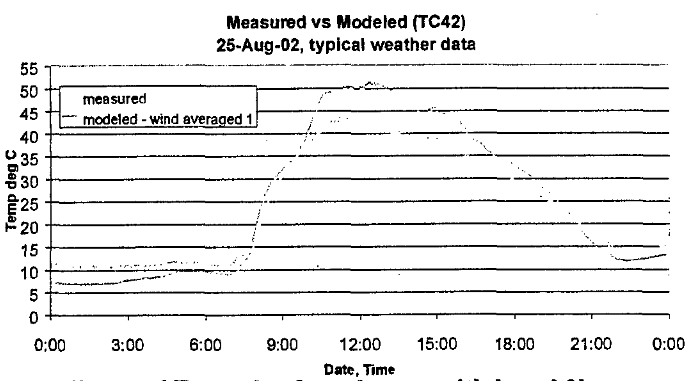

Figure 6 Example of results run with h and 2h.

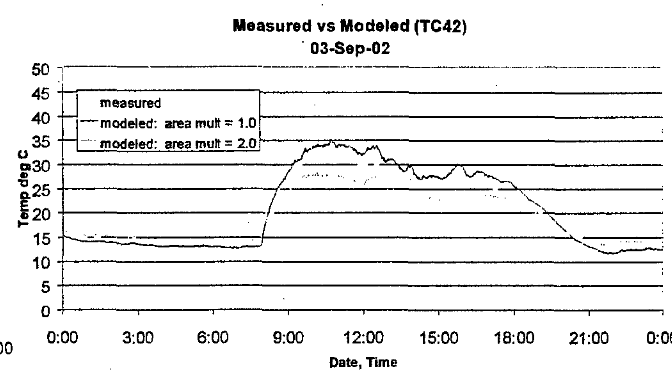

Figure 7 Example Cubi comparison in second trial

Problems with Wind

Ironically, in this experiment (which was carried out in an open field and therefore much more prone to wind) the comparison was worse than the less rigorous previous test. MuSES continually over predicted during the daytime solar loading period and under predicted at night. Based on ongoing discussions, experimenters began to suspect that wind convection was the problem. The convective approach used in MuSES had passed the V&V process up to this point, because it uses accepted textbook theory on the value of the convective coefficient h in wind and the associated Reynolds numbers. Those with experience in this area understand there are problems with this accepted theory. Miller's Law is the source behind the problem, because natural turbulent flows is one of physics unsolved mysteries at this juncture and theory is based on measured data[5]. Any comparison of measured to modeled data will be "unfair" where turbulent wind is a factor, because there is an inherent approximation or error built into the theory. This is because measuring wind out in nature is difficult and therefore theories are developed based on the more predictable continuous flow wind tunnels that can be measured more effectively. But we also know that due to the turbulence created by high speed fluctuations of natural wind, the real value of h (and hence, the effective cooling) is higher than the smooth wind tunnel flows. There is however, no exact theory for calculating the exact speed at which wind moves from laminar to turbulent nor the exact Reynolds number to use to calculate h once it does. It can be as much as 2-3 times more than the accepted textbook value. Using this example, it is perfectly "valid" to understand that using accepted theories, predicted values will result in temperatures that are higher than seen in the field. However, outside of implementing an exhaustive computational fluid dynamics capability within MuSES, we wish to allow the user to get a value more in line with what would be expected in the field. We have determined that we need a better approach. Firstly, to validate that this was indeed the issue, trial and error was used to determine if there could be found a "right" value of h to fit the data. Figure 7 shows the result of standard h and doubling the value of h. The result is that the model now has reversed the trend--under predicting during the day and over predicting during the night. Clearly, in this case, 2h is over compensating. The actual value in this example was a little over 1.5. The next step in validating that this was the problem with the matching was to design an experiment to eliminate the effects of wind and see if this removed the problem. A Plexiglas container was built (figure 5) and placed over a flat plate with a solar sensor and an aspirated air sensor placed inside. Some example results are shown below.

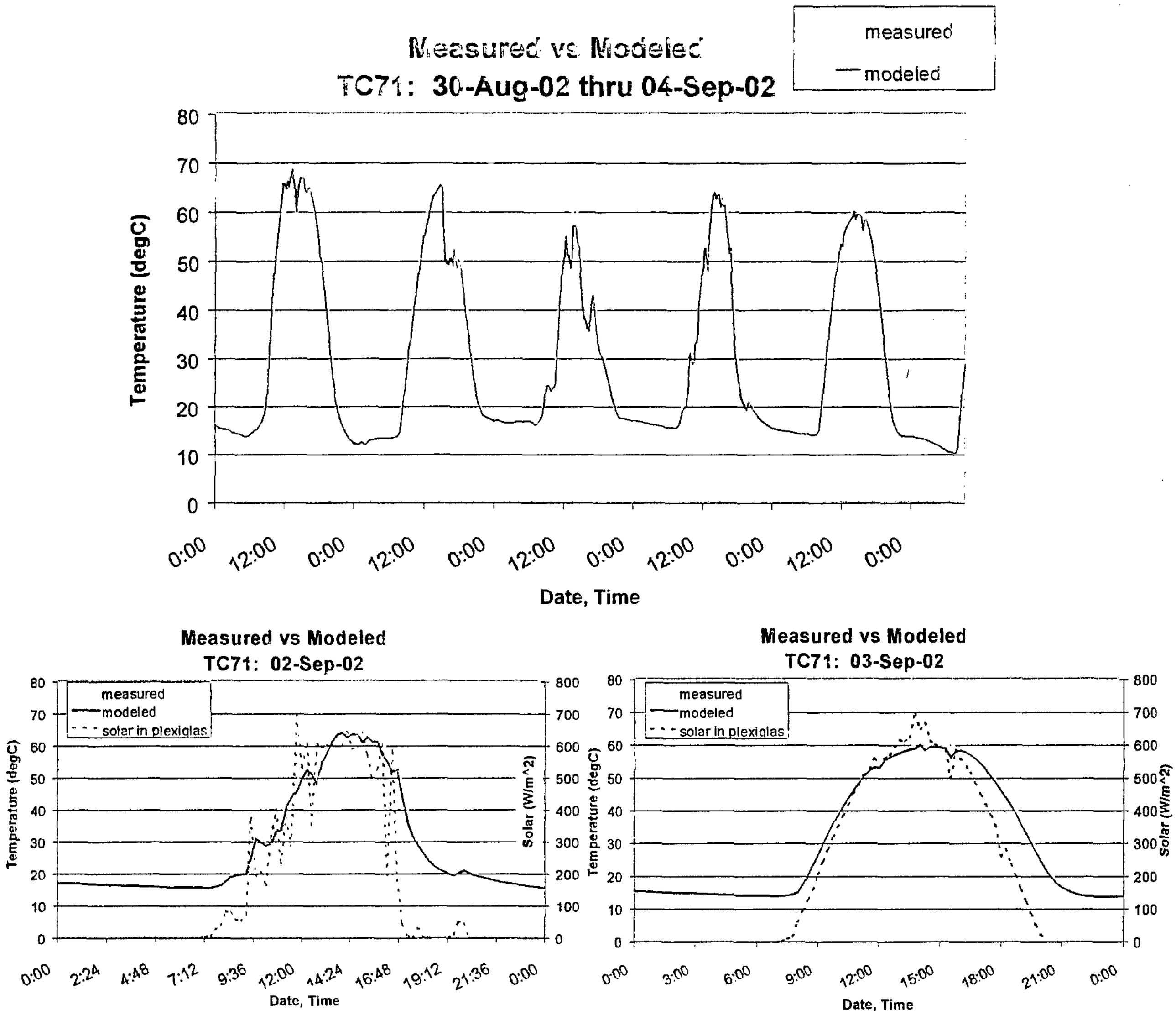

Figure 8 Output from Summer 2002 Covered Plate Trial (a-upper, b-lower left, c-lower right)

This sample is the worst of the data and yet clearly the agreement is quite good--supporting the earlier conclusion. Figure 8a shows five straight days of measurements, while 8b and 8c are close-ups of particular days. There is a small error inherent in the experiment at between 1500h and 1600h, where the solar sensor's view is blocked by corner in the Plexiglas container, effecting the modeled results. Notice the absolute agreement during non-solar hours. In addition, there is some departure during times of fluctuating cloudiness (where cloud solar scatter affects the validation experiments). To aid in analysis, the measured solar data is also plotted to track cloud activity and the blocked solar sensor. These runs were executed with the more simple weather file carried over from PRISM that uses a cloud cover number that is an approximation by an observer.

Additional runs will be made using a more rigorous weather scenario, but for now, the conclusion is clear. The theory might be acceptable, but in order to provide the user with a more robust prediction in nature, an alternative approach will have to be developed for inclusion in MuSES 7.0 (See Lessons Learned Section)

Apparent Temperature

In addition to physical temperatures, validation tests are scheduled for apparent temperatures as well. Figure 9 shows some data from a previous experiment where MuSES predictions of apparent temperature were compared against data taken from imagery. Taking into account temporal trends (how close the data points are to the line, not just in the y-axis) that are much easier to see in the plot on the right, the data falls within the aforementioned 2 degrees C except for one point. The standard deviation also falls within this range and the root mean square (RMS) value is 1. The one-off larger deviation was accounted for in the experiment. More exacting apparent temperature experiments will be conducted using Cubi.

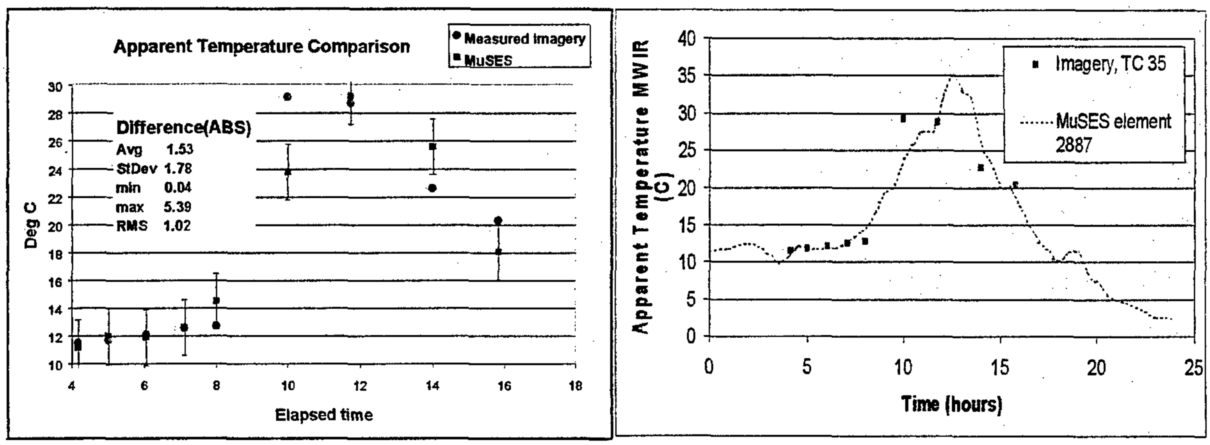

Figure 9 Apparent temperature comparison

LESSONS LEARNED

During the initial test and the second test, many lessons were learned by both persons new to MuSES and by experienced professionals. Part of our V&V process included having new users exposed to the output validation process to investigate problem areas not just with MuSES, but with the modeling experience and in the process of comparing measured and modeled results. Some of the lessons learned were obvious some were not.

New Wind Model

One outcome of the second test is that MuSES 7.0 will have a new wind model. In addition to accepted theory that is known to have its limits, the user will be given a choice of additional approaches. A new model will be developed using the data from the second Cubi/plates experiment. This high quality data, while limited in geometry shape, provides a great deal of new information on flat plates facing in different directions in multiple wind speed situations. This new model will be derived using time of day, plate angle, wind direction and speed and will be compared with the accepted theoretical formula and work of previous researchers.

Material Properties

The materials database in MuSES should be understood to contain <u>representative</u> values for listed items. Depending on the question being asked of the model, this choice might be appropriate. In a validation experiment or if the answer needs to be exact. one has to obtain measurements of the properties. For a validation experiment, one has to know the exact emissivity and absorptivity of the surface of the test object and the measurements need to be made on a coupon that is in the same condition as the validation object measured in the field. Sensitivity analyses show that this area is critical. Lesson: If absolute answers are important it is worth the time or expense to get coupons made and measured. In addition, if the trials potentially will last some time out doors, have coupons measured before the test begins and then keep a coupon with the test set up, so measurements can be made at the end of the trials to check for changes.

Diligent Monitoring of Meteorological Data

It is crucial that the test and measurement equipment be set up properly. The weather station used in the early Cubi trials measured both the shadow band diffuse solar radiance and the total solar radiance. Upon examination, it was determined that the equipment used to measure diffuse solar was allowed to get out of alignment, causing contamination of the measurements by direct sunlight in the morning hours. Lesson: It is critical to pay close attention to this area. In this second phase, a person was put solely in charge of developing, calibrating, and maintaining the met station as well as data reduction. Bad meteorological data can invalidate days of data.

Thermocouple Data

Not only is it crucial that test and measurement equipment be set up properly, but it is also necessary to build in redundancy to maintain good data and mitigate risk against malfunctioning equipment whenever possible. Towards this end, five thermocouples were applied to each face of Cubi. This was planned to help capture gradients along the faces, but more importantly to provide redundancy. This proved valuable, since having more than one thermocouple on each face made it very easy to discover that thermocouple 14 had malfunctioned during the first Cubi trials.

Invalid comparisons of modeled and measured.

A common problem, particularly with inexperienced modelers, is comparing modeled and measured data, where not all the relevant information is put into the model. Often in an attempt to create short cuts, certain assumptions are made to save time. For example, it was determined that in the first Cubi test the plywood base (which extended out past the bottom of Cubi by approximately an inch) actually interacted enough with the model to make a difference. The test location itself was also questions--there was an uncertainty as to whether buildings and trailers in the test vicinity had much of an impact on the test results. This exact problem occurred in Israel's early Cubi tests as well. Comparisons to measured and modeled data were problematic, but the test setup was on a rooftop with building and instrumentation surrounding the test object--these interactions were not included in the model. Lesson: Pay close attention to reduce interactions with items that are not part of the experiment or take the time to model them accurately. Assumptions of non-interaction are often incorrect.

Cloud Scattering Spikes

When taking validation weather data, in is important to be cautious of partly cloudy days and the solar measurement readings. Per Dr. Ann Webb at UMIST, "The phenomenon is well known amongst those who routinely measure solar radiation, and is apparent when measuring both total (all wavelength) radiation and discrete spectral. Radiation travelling thorough the atmosphere is scattered whenever a photon happens to meet a particle: some photons (direct beam radiation) reach the earth uninterrupted. The scattering falls into a number of regimes. Clean air particles provide Rayleigh scattering that goes half in the forward (earth) direction and half backwards (to space). Other constituents of the atmosphere (pollution, cloud droplets) are much bigger and their ratio to the size of the wavelengths of solar radiation put them in the Mie scattering regime. Mie scattering patterns change with the ratio of wavelength to particle size, but generally produce a stronger forward scattering compared to the Rayleigh case. As a cloud comes close to the path of the direct solar beam and circumsolar radiation, this strong region of radiation is subjected to forward scattering ... and an increase or spike in the data record is recorded that can be substantially more than the clear sky value. Usually the cloud then passes in front of the sun and the direct radiation is blocked, causing a significant drop in signal. As the cloud clears the solar disk a further high radiation value is briefly recorded. The net effect of all this is that cloud does decrease radiation reaching the earth (as one would expect), but for transient moments it can increase values above those of clear skies, especially on summer days with scattered fair weather cumulus ("cotton wool clouds")."[6]

As these examples show it is the user's responsibility to determine the appropriate level of accuracy required in any model and simulation and when the resolution is "good enough". This is why we are running sensitivity analyses on the code. In this way, modelers can better determine how much attention needs to be paid to individual parameters in order to achieve predictions within a certain tolerance.

SUMMARY

MuSES/RadTherm is undergoing a great deal of scrutiny by both industry and the government. MuSES 6.0 is proving out to be very robust. While not technically a bug in V&V terms, the wind model needs improvement and there is already a change set for version 7.0 to be released in the late Fall of 2002. The "comparison with measured data" tests, on which many people hang their hats, have proven successful to this point. Additional apparent temperature trials will be conducted using Cubi and V&V will be an ongoing process with each new release--scheduled every six months.

As is the case with other sophisticated engineering level software tools, users need to receive training in the software and need to have at least an intermediate understanding of heat transfer in order to best utilize the code. And as is the case with all models, the user must know what question he or she wants the model to answer and if the model is capable of answering it. At this stage, (i.e. within the bounds of what has been tested) MuSES/RadTherm 6.0 performs within acceptable tolerances and gives expected results.

Unless a modeler is doing a validation or a model-test-model approach, comparison to measured data will most likely not be possible. It is our hope that by these lessons learned and the data contained within, that modelers may better be able to determine the most effective method of modeling for a particular situation.

ENDNOTES AND ADDITIONAL REFERENCES

[1] DMSO "Verification, Validation and Accreditation (VV&A) Recommended Practices Guide, 1996, https://www.dmso.mil/public/transition/vva/; current guidance: https://www.dmso.mil/public/transition/vva/policiesguidance; plus the updated DoD instructions: https://www.dmso.mil/public/library/projects/vva/products/DoDI%205000.61%20-dbl%20spa%20draft_7-10-2002a.doc

[2] http://www.amso.army.mil/; http://www.amso.army.mil/library/index.htm

[3] DMSO "Verification, Validation and Accreditation (VV&A) Recommended Practices Guide, 1996

[4] Brian Miller, research engineer, CECOM, Night Vision Electronics and Sensors Directorate (NVESD).

[5] Mills, A. F., Basic Heat and Mass Transfer 2/E, 1999 Prentice Hall, Inc., p382

[6] Per email, Dr Ann Webb, a solar researcher at UMIST in England.

Additional References:

http://www.informs-cs.org/

http://www.acm.org/pubs/tomacs/

http://www.sisostds.org/